**📊 Actividad práctica — Introducción a la Ciencia de Datos**

**1. Definición y evolución de la ciencia de datos**

**Línea de tiempo (5 hitos clave):**

1. **Siglo XIX – Estadística tradicional:** Uso de métodos matemáticos para describir fenómenos sociales y económicos.
2. **Década de 1960 – Bases de datos relacionales:** Estructuración y almacenamiento eficiente de grandes volúmenes de información.
3. **Década de 1990 – Minería de datos:** Descubrimiento de patrones útiles en conjuntos de datos mediante algoritmos.
4. **Década de 2000 – Big Data:** Aparición de tecnologías para procesar datos masivos y no estructurados en tiempo real.
5. **2010 en adelante – Ciencia de datos e IA:** Integración de estadística, programación y machine learning para predecir, automatizar y tomar decisiones inteligentes.

**Importancia en el contexto actual:**  
La evolución de la ciencia de datos ha permitido transformar los datos en un activo estratégico. Hoy en día, desde startups hasta gobiernos utilizan IA y Big Data para optimizar procesos, anticipar tendencias y personalizar servicios. En un mundo hiperconectado, quien maneja los datos tiene la ventaja competitiva.

**2. Perfil y competencias del científico de datos**

**Competencias técnicas:**

1. Programación en Python/R/SQL
2. Machine Learning y Deep Learning
3. Manejo de bases de datos
4. Visualización de datos (Tableau, Power BI, Matplotlib)
5. Cloud computing y Big Data (Hadoop, Spark)

**Competencias blandas:**

1. Pensamiento crítico
2. Comunicación efectiva
3. Resolución de problemas
4. Trabajo en equipo multidisciplinario
5. Adaptabilidad al cambio

**Relación con sectores:**  
Estas competencias permiten, por ejemplo, en **salud** mejorar diagnósticos predictivos; en **finanzas**, detectar fraudes; en **marketing**, segmentar clientes con precisión; y en **gobierno**, optimizar recursos y políticas públicas. En todos los casos, el científico de datos actúa como traductor entre el lenguaje técnico y la toma de decisiones estratégicas.

**3. Diferencias entre estadística, minería de datos y ciencia de datos**

| **Disciplina** | **Definición** | **Herramientas** | **Objetivos** | **Ejemplo de aplicación** |
| --- | --- | --- | --- | --- |
| **Estadística** | Rama matemática que analiza y resume datos. | SPSS, R, Excel | Describir y explicar fenómenos. | Encuestas de opinión. |
| **Minería de datos** | Extracción de patrones ocultos en grandes volúmenes de datos. | Weka, RapidMiner, SQL | Encontrar relaciones y tendencias. | Segmentación de clientes en un banco. |
| **Ciencia de datos** ⭐ | Disciplina interdisciplinaria que integra estadística, informática e IA. | Python, TensorFlow, Spark, Power BI | Predecir, optimizar y automatizar decisiones. | Recomendaciones de Netflix. |

👉 **Diferencia clave:** La estadística describe, la minería descubre, y la ciencia de datos **predice y crea valor estratégico**.

**4. Ciclo de vida de un proyecto de ciencia de datos (CRISP-DM)**

**Fases:**

1. Comprensión del negocio
2. Comprensión de los datos
3. Preparación de los datos
4. Modelado
5. Evaluación
6. Despliegue

**Ejemplo:**  
En un supermercado, se recolectan datos de compras de clientes. Se preparan y limpian los datos, se entrena un modelo de segmentación de clientes, se evalúan los resultados, y finalmente se despliega un sistema que recomienda promociones personalizadas según el historial de compra.

**Diagrama (PlantUML):**

@startuml

!theme spacelab

title CRISP-DM - Ciclo de Vida

start

:Comprensión del negocio;

:Comprensión de los datos;

:Preparación de los datos;

:Modelado;

:Evaluación;

:Despliegue;

stop

@enduml



**5. Casos de éxito y aplicaciones actuales**

**Caso 1 – Netflix (Recomendaciones personalizadas)**

* **Problema inicial:** Los usuarios abandonaban la plataforma al no encontrar fácilmente contenido de su interés.
* **Aplicación de ciencia de datos:** Implementación de algoritmos de machine learning que analizan patrones de consumo y recomiendan películas y series.
* **Impacto:** Incremento de retención de usuarios y reducción de cancelaciones (churn). Netflix atribuye más del 75% de las visualizaciones a su sistema de recomendación.

**Caso 2 – Waze (Optimización de rutas de tráfico)**

* **Problema inicial:** Conductores perdían tiempo por tráfico impredecible y rutas ineficientes.
* **Aplicación de ciencia de datos:** Uso de datos en tiempo real enviados por millones de usuarios para predecir atascos y proponer rutas alternativas.
* **Impacto:** Millones de personas ahorran tiempo y combustible diariamente; además, gobiernos utilizan sus datos para mejorar la planificación vial.